Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Relative Liquidity and Future Volatility

Marcela Valenzuela, Ilknur Zer, Piotr Fryzlewicz, and Thorsten Rheinlander

2014-45

Relative Liquidity and Future Volatility*

Marcela Valenzuela[†] Ilknur Zer[‡]
Piotr Fryzlewicz[§] Thorsten Rheinlander[¶]

May 2014

Abstract

The main contribution of this paper is to identify the strong predictive power of the relative concentration of depth provision, rather than volume of orders, over volatility. To this end, we propose a new measure, relative liquidity (RLIQ), which extracts information from a limit order book distribution and captures the level of consensus on a security's trading price. Higher liquidity provision farther away from the best quotes, relative to the rest of the book, is associated with a disagreement on the current price and followed by high volatility. The relationship is robust to the inclusion of several alternative measures.

Keywords: order-driven markets, limit order book distribution, volatility predictability, liquidity

JEL Classification: G1, G20

*We sincerely thank Vicente Cunat, Jon Danielsson, Christian Julliard, and Richard Payne for their valuable comments. We would like to thank to Coskun Gunduz, Recep Bildik, and Huseyin Eskici for providing us data and the support for understanding the market mechanisms.

[†]University of Chile, Department of Industrial Engineering, Republica 701, Santiago, Chile.

[‡]Federal Reserve Board, 20th Street and Constitution Avenue N.W. Washington, D.C. 20551, USA. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

[§]London School of Economics, Department of Statistics, Houghton Street, London, WC2A 2AE, UK.

[¶]Vienna University of Technology, Research Group Financial and Actuarial Mathematics (FAM), Wiedner Hauptstrasse 8-10/105-1, A-1040, Vienna, Austria.

1 Introduction

This paper examines the link between two central concepts in financial markets: liquidity and volatility. Liquidity, the ease with which an asset can be traded without affecting the asset's price, is essential for well-functioning financial markets. Hence, understanding the effects of liquidity provision on market dynamics has gained an increased attention from regulators, market participants, and academics alike. On the other hand, information on volatility, variation in trade prices, is one of the main ingredients in assessing risk-return trade-off for portfolio valuation, derivatives pricing models, and it is important for the calibration of execution probability of limit orders.

Liquidity manifests itself in three dimensions: tightness, i.e., bid-ask spread; depth, which is a measure of price impact; and resilience which is related to the speed of price reversals (Porter (2008)). When the limit order book is thin, i.e., when the volume of orders available to trade at the best quotes is low, any market order has a price impact which will translate into higher volatility. This could be viewed as a "mechanical" liquidity–volatility link. On the other hand, the relative concentration of depth provision at each quote reveals information of the disagreement on the true price. According to the theoretical predictions of Goettler, Parlour, and Rajan (2005, 2009) higher liquidity provision around the best quotes relative to the rest of the book is associated with a consensus on the current price; whereas the accumulation of orders at a quote farther away from the best prices signals to the market that current quotes are mispriced. We argue that, in the latter case price movements are more plausible, creating higher future volatility. In this paper, we construct a measure of relative concentration of depth provision in a limit order book and study its informativeness on future volatility.

The proposed measure, *relative liquidity* (*RLIQ*), is the first principal component of the distribution of orders waiting at the aggregate limit order book. To calculate RLIQ, we first obtain the empirical probability density function of a limit order book for a given stock. We then calculate the cross-sectional average of individual stock distributions to reach the aggregate distribution. In other words, we measure the proportions of orders waiting at each price level in the market. Finally, we employ principal component analysis to summarize this information in as few interpretable quantities as possible.

Many studies examine the effects of the volume of orders on the trading decisions of agents (Parlour (1998), Ranaldo (2004), Foucault, Kadan, and Kandel (2005), Ellul, Holden, Jain, and Jennings (2007), Cao, Hansch, and Wang (2008, 2009), and Pascual and Veredas (2009), among others). The main conclusion we extract from these studies is that depths at and farther away from the best quotes play different roles in traders' order choices. Hence, as a summary measure of a limit order book, RLIQ has three ingredients: first it summarizes the *relative* distribution of orders waiting to be traded, which reveals information on the disagreement of the true price, second it considers the information contained beyond the best quotes, and finally it weighs this information based on price distances. The weighting scheme captures the different levels of informativeness of the quotes. Instead of imposing an exogenous weighting scheme, we employ the loadings of the first principal component of the aggregate distribution to weigh the information provided by different price levels. Thus the principal component analysis enables us to avoid the subjective judgments regarding the relative importance of quotes.

We evaluate the predictive power of relative liquidity over both market and individual stock volatilities at an intraday level, with a particular interest in the former. Predicting market volatility is important because it can be used by policy makers as a proxy to the vulnerability of financial markets, as changes in market volatility have systemic repercussions on the whole economy (Schwert (1989) and Poon and Granger (2003)). During stressed market conditions, liquidity may disappear very quickly. For example, the withdrawal of the high-frequency liquidity providers has contributed to the volatility presented in the flash crash of 2010 within minutes (Kirilenko, Kyle, Samadi, and Tuzun (2011)). This makes it desirable to study the volatility–liquidity relationship at an intraday level.

Forecasting volatility in a high-frequency setting has important implications on traders' order choice strategies. Placing a buy (sell) limit order is equivalent to writing a free put (call) option to the market (Handa and Schwartz (1996)). The higher the volatility, the higher the option value of the limit order, since in this case the probability that the spot price hits the limit price increases. There is extensive evidence, both theoretical and empirical, that investors submit limit orders in high volatility states (see Foucault (1999) and Ranaldo (2004) for instance) and reduce the execution cost when volatility is expected to be high.

The order and trade books of the largest 30 stocks from the Istanbul Stock Exchange (ISE) form the dataset that we use in this study. By matching these two books and removing the executed orders, we dynamically reconstruct the limit order book. That is, for a given time we have the best bid and ask prices, all of the orders waiting to be executed, their submitted prices, and their corresponding volumes. Since ISE is a fully centralized purely order driven market and operates with a single trading platform, our data contains the entire order flow in the public domain, which brings a major advantage compared to the main European and U.S. market exchanges.

Pascual and Veredas (2010) separate transitory volatility from informational volatility by employing a dynamic state-space cointegration model for bid and ask quotes. By focusing on assets with large tick sizes, Delattre, Robert, and Rosenbaum (2013) propose a statistical methodology to estimate the efficient price of an asset through the order flow. In this paper, we proxy the volatility of the true price process by employing the two scales realized volatility estimator proposed by Zhang, Mykland, and Ait-Sahalia (2005) and Ait-Sahalia, Mykland, and Zhang (2011), which gives an unbiased and consistent estimate of the quadratic variation of the true price process.

We show that relative liquidity (RLIQ) is the strongest among standard liquidity and trading activity measures, in explaining the in-sample variations in market volatility. On average, one standard deviation increase in RLIQ decreases the 15-minutes-ahead volatility by 4.4 basis points. Finally, we examine both the absolute and the relative volume of orders and document that the predictive power of absolute depth is no longer significant under the presence of RLIQ.

Out-of-sample forecasting tests provide evidence for substantial forecasting power of relative liquidity. It predicts 15-minutes-ahead market volatility at a 5% level with an out-of-sample R^2 of 12.9%, where the forecasting power lasts up to 75 minutes ahead. We also document that capturing both the relative liquidity and the tightness dimension of liquidity delivers an out-of-sample R^2 of over 24%. Finally, we show that the time-series relationship between RLIQ and market volatility is not driven by variations in a particular stock or industry, but rather that it is shared by the majority of the stocks. We find a significant relationship between the individual stock level RLIQ and future volatility for 87% of the stocks in our sample.

This paper relates to the literature that attempts to measure the liquidity provision considering the whole book. Domowitz, Hansch, and Wang (2005) propose an illiquidity measure based on the supply and demand step functions and conclude that the stock correlated movements in liquidity, i.e., the liquidity commonality, is priced in stock returns. Marshall (2006) defines liquidity by the weighted order value, which depends on the execution rate of orders waiting in each price band and their corresponding price and volumes. The author documents a negative association between liquidity and monthly returns. In another related study, Naes and Skjeltorp (2006) introduce the slope of the book, which describes the average elasticity across all price levels with the corresponding volumes.

This paper is part of the market microstructure literature that examines the predictive power of liquidity on intraday volatility. In an early empirical work, Ahn, Bae, and Chan (2001) analyze the interactions between transitory volatility and order flow composition. They conclude that the transitory volatility arises mainly from the scarcity of limit orders at the best quotes. Pascual and Veredas (2010) show that trade size and quoted depth both at the best and away from the quotes have a predictive power for individual volatility. Duong and Kalev (2008) investigate the forecasting power of the Naes and Skjeltorp (2006)'s definition of order book slope. By using data from the automated futures market, Coppejans, Domowitz, and Madhavan (2001) study the dynamic relationship between liquidity, return, and volatility in a vector autoregressive framework.

Finally, the paper is related to the few studies that use intraday data from the Istanbul Stock Exchange (ISE). Ekinci (2008) and Koksal (2012) provide descriptive analyses of the intraday liquidity patterns of the ISE by focusing on the behavior of spreads, depths, and trading volume. Using different stocks and time periods, both studies conclude that the liquidity related variables follow the usual U-shape pattern for both of the trading sessions. Valenzuela and Zer (2013) study how the market characteristics and information content of the limit order book affect the order choice of investors.

Our contribution to the literature is threefold: first, we provide a new variable that summarizes the information provided by a limit order book. Contrary to the aforementioned measures, which focus on the absolute volume of the orders waiting in a given book, RLIQ is based on the relative distribution of volume at a given time. Second, we show that relative liquidity contains information on future volatility that cannot be explained by the standard predictors of volatility. Finally, in contrast to these former studies, which examine the volatility–liquidity relationship at an individual stock level, we focus on the link between aggregate liquidity and future market volatility.

The rest of the paper is organized as follows: the next section describes data and the trading structure in our market. Section 3 explains the construction of our measure in detail. Section 4 introduces the econometric methodology and variables included in the analysis. The in-sample and out-of-sample predictive results, and robustness checks are given in Section 5. Finally, Section 6 concludes.

2 The Market and Data

Our dataset comprises order and trade books of the individual constituents of the Istanbul Stock Exchange ISE–30 index for the period of June and July 2008. The index corresponds to almost 75% of the total trading volume of the ISE for the sample period. The ISE is a fully computerized as well as a fully centralized purely order driven stock exchange, i.e., the trading of the listed stocks has to be executed in the ISE via electronic order submissions without a market maker. Hence, our data fully captures the order flow.

The trading occurs between 09:30am and 5:00pm, with a lunch break. Similar to all other major exchanges, a trading day starts with a call market matching mechanism of 15 minutes to determine the opening price. In contrast to the opening session, during the continuous double auction, all of the orders submitted are either matched instantaneously based on the usual price and time priorities or booked until the corresponding match order arrives to the system. A submitted order is valid for a given session or for a day. All brokers have access to the full book. Prior to the submission of an order, they can see the quantity available at different prices, not limited to the best five or ten quotes.

Order book data consists of information regarding the orders submitted for a given stock, whereas trade data records the executed orders, both time-stamped at the accuracy of 1 second. The order and trade ID numbers generated by the exchange system allow us to identify the priority of orders submitted in the same second, to match orders in the order and trade books, and finally to track any order through submission to (possible) execution or modification. By using the order and trade books, we first reconstruct the limit order book dynamically for each stock and obtain relevant information, such as the bid and ask prices and corresponding volumes at a given time. Hence, the reconstruction methodology enables us to obtain snapshots of a limit order book at any given time. In particular, we have the same information that a trader observes: the volume of orders waiting to be executed for the entire price range. We use this information to calculate the relative frequency of orders waiting in every price level.

3 The Limit Order Book Distribution and Relative Liquidity

3.1 The Limit Order Book Distribution

We obtain the limit order book distribution by employing the following steps, which are illustrated with an example in Appendix A:

1. For each security and each day, we sample the limit order books every 15 minutes, excluding the lunch break and the opening session. The first snapshot of the book contains the unexecuted orders submitted until 10:00, whereas the last one contains all of the unexecuted orders submitted until 17:00. Hereafter, the time subscript τ indexes these trading intervals, with $\tau = 1, 2, ..., 21$. We repeat the empirical analysis with 30–minute sampling frequencies as a robustness. The results are presented in Section 5.6.

2. We calculate the (tick-adjusted) price distance of each limit order relative to the best limit price in each snapshot. In other words, for each order i in the limit order book at τ, we define the price distance Δ as:

$$\begin{aligned} \Delta_{i,\tau}^{\text{buy}} &= (p_\tau^B - p_i^{\text{buy}})/\text{tick}, \\ \Delta_{i,\tau}^{\text{sell}} &= (p_i^{\text{sell}} - p_\tau^A)/\text{tick}, \end{aligned}$$

   where p_τ^B (p_τ^A) is the best bid (ask) price in interval τ and p_i^{buy} (p_i^{sell}) is the limit price of the i^{th} order.

3. For each side of the book, day, and limit order book at τ, we get the limit order book probability density function (LOB–PDF) by calculating the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, .., \Delta_c$, where Δ_c is the maximum price distance considered. Therefore, LOB–PDF summarizes both the relative magnitude of the depth provision and its price location.

4. We calculate the equally-weighted cross-sectional averages of individual LOB–PDFs to obtain the aggregate LOB–PDF (avgPDF). In other words, we take the average of the

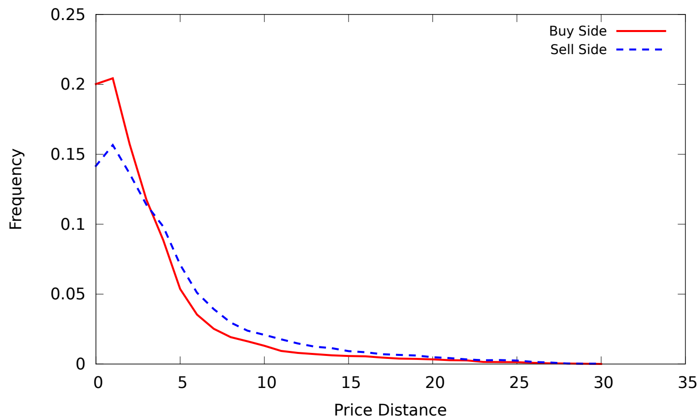


Figure 1: Limit order book probability density function (LOB–PDF) for the period of June and July 2008, averaged across thirty stocks, twenty-one 15-minutes trading intervals, and thirty-nine days considering the whole book.

frequencies in each corresponding bins of the 30 individual stock LOB–PDFs. That represents the proportion of orders waiting at each price level in the market. For a given trading interval τ and price distance $\Delta = 0, 1, 2, .., \Delta_{\mathrm{c}}$,

$$\mathrm{avgPDF}^{\mathrm{buy}}_{\Delta,\tau} = \frac{1}{S}\sum_{s=1}^{S} f^{\mathrm{buy}}_{s,\tau}(\Delta), \tag{1}$$

where $f^{\mathrm{buy}}_{s,\tau}(\Delta)$ is the buy side LOB–PDF of stock s and S is the total number of stocks. The measure for the sell side is calculated analogously. In order to consider the possible impact of bigger or more actively traded stocks, we also calculate the value-weighted and number-of-trades-weighted averages of the individual LOB–PDFs to obtain the avgPDF. We reach qualitatively similar results. Discussions are presented in Section 5.6.

Figure 1 reveals that for both sides of the market, the frequency of orders submitted at the second best quotes is the highest and the limit order book distribution is positively skewed. The liquidity provision is concentrated closer to the best quotes for the buy side compared to the sell side, which can be observed by comparing either the mean or the skewness of the distribution presented in Table 1. The mean of the distribution, for all of the time intervals, is higher for the sell side than the buy side. Wilcoxon rank sum test concludes that the difference is statistically significant at a 5% level. Around 40% and 30% of the depth is concentrated at the best or second best quotes ($\Delta = 0$ or $\Delta = 1$) for buy and sell sides, respectively. The cumulative frequency of orders waiting 5 or more ticks away from the quotes is 35% for the sell side, whereas it is only 23% for the buy side. The Kolmogorov-Smirnov test rejects the hypothesis that the buy-side and sell-side distributions are equal, at a 1% significance level. Finally, the average variance of the sell side is 36% higher than the average variance of the buy side, indicating that the buy side is less dispersed.

Around 90% of the submitted orders are waiting within the 10 best prices for both sides of the market. Hence, we only consider the information contained in the book up to the 10^{th} best quotes. In other words, the main discussions presented in the rest of the paper are obtained by setting $\Delta_{\text{c}} = 10$. However, we examine the robustness of our findings when Δ_{c} is equal to 20 and 30, i.e., when we consider the whole book, in Section 5.6.

3.2 Summarizing the limit order book distribution: RLIQ

The shape of the limit order book distribution at time τ is given by the proportion of volume waiting to be traded at different price distances Δ. There are several ways to summarize this information. We want our summary measure to weigh the information provided in different quotes based on price distances to capture the different levels of informativeness of the quotes. One, for example, could assume exogenously given weights or give equal weights to the frequency of orders waiting at each price distance Δ. We instead employ the principal component analysis (PCA), which extracts the most important (uncorrelated) sources of variation in the LOB distribution. The advantage of this approach is that, it assigns an objective weighting scheme, which aims to encode as much information about the LOB distribution in as few quantities (principal components) as possible.

The principal component analysis applied on the ten bins of avgPDF defined in (1) produces ten uncorrelated principal components. The first principal component is the leading eigenvector in the spectral decomposition of the covariance matrix of avgPDF, which explains the highest variation in the limit order book distribution. Our relative liquidity summary measures, $\mathrm{RLIQ}^{\mathrm{buy}}$ and $\mathrm{RLIQ}^{\mathrm{sell}}$, are chosen to be the first principal components of avgPDF for both sides of the market. In Section 5.6, we discuss the sensitivity of the findings by (a) considering the first three and five principal components, (b) using the empirical frequencies of orders waiting at each price distance separately, and (c) employing LASSO (Tibshirani, 1996), a shrinkage and variable selection technique.

Figure 2 plots the loadings of $\mathrm{RLIQ}^{\mathrm{buy}}$ and $\mathrm{RLIQ}^{\mathrm{sell}}$ that are used to weigh the information provided by different price levels. The signs of the loadings of the first principal component are chosen so that the sign corresponding to first price distance ($\Delta = 0$) is positive, both for the buy and sell sides of the market. The figure reveals that if the frequency of orders waiting around the best quotes increases, this increases RLIQ due to positive loadings assigned to the information provided by the top of the book. On the other hand, an increase in the proportion of orders waiting farther away from the best quotes translates into a decrease in RLIQ. This easy-to-interpret pattern in the loadings provides further justification for the use of this summary measure as a possible predictor of volatility.

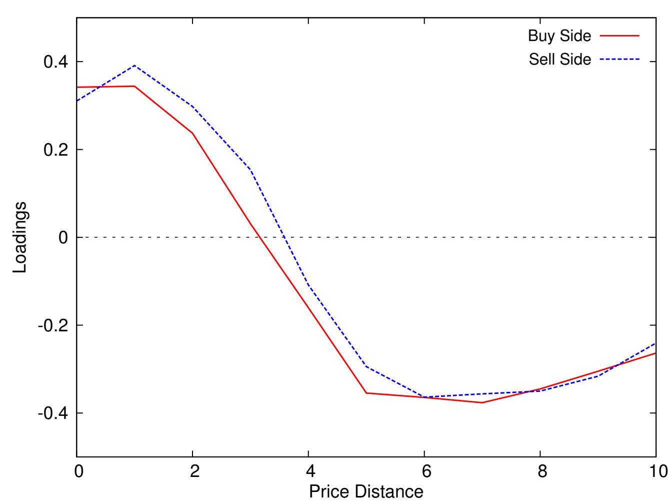


Figure 2: Loadings of the first principal component of the aggregate limit order book distribution (avgPDF) defined in (1), when $\Delta_c = 10$.

4 Predictive Analysis

4.1 Methodology

To evaluate the information content of a limit order book on future volatility, we rely on a standard predictive regression model of intraday volatility:

$$\begin{aligned}\sigma^M_{\tau+1} &= a_0 + a_1\sigma^M_\tau + a_2\,\mathrm{RLIQ}^{\mathrm{buy}}_\tau + a_3\,\mathrm{RLIQ}^{\mathrm{sell}}_\tau + \sum_{k=1}^{20} b_k T_{k,\tau} \\ &+ \ \mathrm{controls} + \varepsilon_\tau,\end{aligned} \tag{2}$$

where for a given interval τ, σ^M_τ is the mid-quote-volatility of the value-weighted index, and $\mathrm{RLIQ}^{\mathrm{buy}}_\tau$ and $\mathrm{RLIQ}^{\mathrm{sell}}_\tau$ are the proposed relative liquidity summary measures, calculated as the first principal component of the aggregate limit order book distribution for buy and sell sides of the market, respectively. $T_{k,\tau}$ is the intraday dummy that equals to 1 if $k = \tau$. We include the lagged volatility, σ^M_τ, and intraday dummies in the set of explanatory variables to control the well-known systematic intraday patterns and clustering in volatility. Furthermore, we employ both the standard predictors of volatility and other liquidity measures as control variables, which are introduced in Section 4.3.

4.2 Measuring volatility: the two scales realized volatility estimator (TSRV)

The topic of volatility forecasting and market microstructure noise, which arises from several sources inherent in the trading process, such as the informational effects, temporary liquidity withdrawals, bid–ask bounces or data recording errors has been examined extensively in the literature. The early approach to estimate the realized variance was from the sum of frequently sampled squared returns. However, this estimation approach is not entirely appropriate under the presence of microstructure noise. Relevant studies have addressed this potential problem and have proposed ways to improve the estimator. Ait-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2009), Ghysels and Sinko (2011), among others focus on optimal sampling frequency. Zhou (1996) and Hansen and Lunde (2006) consider a first-order autocorrelation to bias-correct the realized variance. Finally, Zhang, Mykland, and Ait-Sahalia (2005) and Ait-Sahalia, Mykland, and Zhang (2011) provide the two scales realized volatility (TSRV) estimator, which is the volatility estimator that we employ in this study.

Let X denote the fundamental log-stock price process. In financial data, instead, we can only observe log-price Y, either in a form of transaction or quoted price, which is often modeled as a linear combination of X and some noise ϵ. The noise term is usually assumed to be i.i.d. and independent from the X process, with X following a geometric Brownian motion (See Ait-Sahalia, Mykland, and Zhang (2005), Zhang, Mykland, and Ait-Sahalia (2005), and Ait-Sahalia, Mykland, and Zhang (2011) for details). In the presence of the microstructure noise, ϵ, the TSRV estimator enables the use of the full available sample data, and gives an unbiased and consistent estimate of the volatility of the true price process. The TSRV is defined as:

$$\langle X, X\rangle_T^{TSRV} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}[Y,Y]_T^{sparse,k} - \frac{1}{K}[Y,Y]_T^{(all)}}, \tag{3}$$

where $[Y,Y]_T^{(all)}$ is the realized variance calculated using the whole sample with size T. To obtain $[Y,Y]^{sparse,k}$, we first divide the whole sample into K moving window subsamples (following Ait-Sahalia, Mykland, and Zhang (2011), K is set be 5 minutes) with a fixed length of N, where $N = T - K$. For example, the first subsample starts with the first and ends with the N^{th} observation, whereas the second subsample starts with the second and ends with $(N+1)^{\text{th}}$ observation. Then, we sample sparsely with 30-seconds frequency. So, $[Y,Y]^{sparse,k}$ is the realized variance estimator of the k^{th} 30-seconds-sampled mid-quote returns.

4.3 Control variables

Our first set of covariates includes the variables that have been shown as predictors of volatility in the current literature. First, consistent with Bollerslev and Domowitz (1993), Jones, Kaul, and Lipson (1994), and Foucault, Moinas, and Theissen (2007), the number of trades occurring in interval τ, NT, and the average trade size, AQ, are included to capture the trading activity. In a related study, Foucault, Moinas, and Theissen (2007) show that the bid-ask spread is informative of future individual stock volatility. Hence, we also include the relative spread, relSPR$_\tau$, which is calculated as the ratio of the bid-ask spread to the mid-quote prices for each interval. Finally, we consider the slope of a limit order book, SLOPE, as an explanatory variable following Naes and Skjeltorp (2006) and Duong and Kalev (2008). SLOPE aggregates the price-quantity information in different quotes and measures the sensitivity of the quantity supplied in the book with respect to the prices.

Our second set of covariates includes other liquidity measures. We first consider standard depth measures. The depth, defined as the total volume available to be traded at the best bid or ask prices, is one of the traditional measures of liquidity. Note that the notion of depth is different from the one used in Porter (2008). We calculate DEPTHi^{buy} (DEPTHi^{sell}) for $i = 1, 2, ..., 5$, which denotes the volume of orders waiting at the i^{th} best bid (ask) to capture the volume available at and beyond the best quotes for the buy and sell sides of the market, respectively. Second, we employ the Amihud (2002)'s illiquidity measure, AMR, which is calculated as the ratio of absolute stock return to the turnover. Another related illiquidity measure is the log quote slope, logQS, which is introduced by Hasbrouck and Seppi (2001). A decrease in the logQS means that the slope of the best quotes is flatter and the market is more liquid. Finally, we consider the illiquidity measure proposed by Domowitz, Hansch, and Wang (2005), DHW. DHW measures the cost of buying and selling Q shares of the stock, simultaneously. Illiquid stocks are associated with lower liquidity. In this paper, we set Q as the median of the accumulated volume of orders waiting in the book for a given stock.

All of the control variables are calculated as the equal-weighted cross-sectional average of the individual stock measures. As a check of robustness, we repeat the analysis by calculating the value-weighted average of the explanatory variables to proxy the aggregate measures. The results are presented in Section 5.6 and we conclude that the main findings are also confirmed in these regressions.

Compared to standard liquidity measures like spread, depth, and ratios based on both spread and depth, global depth provides a more complete picture of the liquidity provision by considering the book beyond the best quotes. Moreover, instead of focusing on the size of the orders waiting, RLIQ is based on the distribution of volume at a given time, which reveals information of the disagreement on the true price.

5 Empirical Findings

5.1 One-period-ahead predictive regressions

The first focus of our analysis is to examine the predictive power of the relative liquidity, RLIQ, for the 15-minutes-ahead market volatility. To account for the intraday patterns all of the specifications include 21 trading intervals as intraday dummies. To conserve space, we do not report the estimated coefficients of the dummy variables. To improve the ease of interpretation of the estimated coefficients, all of the explanatory variables are standardized to have mean zero and unit variance, and the dependent variable is presented in percentage terms. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and presented in parentheses.

Table 2 reveals that RLIQ variables are significantly and negatively related to the one period ahead market volatility at a 5% level. The proportion of the variation in market volatility explained by our measures are almost 22%. When the intraday dummies are not included in the specification, $\text{RLIQ}_\tau^{\text{buy}}$ and $\text{RLIQ}_\tau^{\text{sell}}$ alone explain around 16% of the variation in volatility. As Figure 2 shows, the loadings of RLIQ are positive for the proportion of volume around the best quotes and turn to negative for the orders waiting away from the best five quotes. Hence, the negative sign of the coefficients indicates that an increase in liquidity beyond the best quotes relative to the top of the book is likely to be followed by a higher level of volatility in the next period. If the volume of orders waiting to be executed is more accumulated beyond the best prices, incoming investors may interpret this as mispricing of the current quotes. Hence, large price movements are more likely to happen creating higher future volatility.

As expected, lagged volatility is highly and positively related to one-period ahead volatility. However, the predictive power of $\text{RLIQ}_{\tau}^{\text{buy}}$ is higher than the lagged volatility. One standard deviation increase in $\text{RLIQ}_{\tau}^{\text{buy}}$ decreases 15–minute ahead volatility by 4.4 basis points, whereas one standard deviation increase in volatility increases the next period volatility by 3.7 basis points. Column III shows that when relative liquidity variables and lagged volatility are included in the specification together, the adjusted R^2 increases to over 25%.

Columns IV and V confirm the robustness of the predictive power of RLIQ for the one-period-ahead market volatility when the standard predictors of volatility and alternative liquidity measures are considered. Not surprisingly, relative spread ($\overline{\text{relSPR}}_{\tau}$) is informative of future volatility at a 5% level. Note that by construction, RLIQ does not include the bid-ask spread since the price distances are calculated as the position to the best quotes, rather than the mid-quotes. Thus RLIQ is related to the depth dimension of liquidity and can be thought as a complement of spread. Moreover, the slope of the book ($\overline{\text{SLOPE}}_{\tau}$), and the slope of the best quotes ($\overline{\text{logQS}}_{\tau}$) are positively and significantly correlated with future volatility. $\text{RLIQ}_{\tau}^{\text{sell}}$, on the other hand, loses its significance when all of the control variables are included in the setting. This result is consistent with the literature documenting that buy orders are more information-driven than sell orders, i.e., the informed traders may exploit their informational advantage by submitting buy orders (see, for instance, Burdett and O'Hara (1987), Griffiths, Smith, Turnbull, and White (2000), and Duong and Kalev (2008), among others). Our results further extend the findings of Duong and Kalev (2008) and Foucault, Moinas, and Theissen (2007), who document that the slope of the book and relative spread, respectively, have explanatory power for future individual stock volatilities. We show that the cross-sectional average of both measures ($\overline{\text{SLOPE}}_{\tau}$ and $\overline{\text{relSPR}}_{\tau}$) have explanatory power for the market volatility as well. Moreover, we provide new empirical evidence that the measure of Hasbrouck and Seppi (2001), log quote slope, is significantly and positively related to the subsequent market volatility. Yet, the estimated (standardized) coefficients and t-statistics of $\mathrm{RLIQ}_{\tau}^{\mathrm{buy}}$ are always the highest among the alternative variables and they are robust and stable in all of the specifications examined. The adjusted R^2 increases from 25.15% to only 29.53% when all of the set of controls are included.

5.2 Relative liquidity vs. standard depth measures

A thin limit order book will be followed by higher volatility since any "large" market order has a price impact. On the other hand, relative liquidity (RLIQ) extracts the relative accumulation of orders rather than the absolute volume of orders in a given book. In other words, a thin book is not necessarily associated with low RLIQ. However as it is constructed from the distribution of volume at a given time, it may share common information with standard depth variables. Thus we next examine whether RLIQ is still significant in explaining subsequent volatility under the presence of depth variables. To this end, we include the volume of orders at different prices along with the RLIQ measures in our analysis. Similarly, all of the specifications include the interval dummies and lagged volatility as control variables.

Table 3 shows that $\overline{\mathrm{DEPTH1}}_{\tau}^{\mathrm{buy}}$ and $\overline{\mathrm{DEPTH1}}_{\tau}^{\mathrm{sell}}$, the total volume of orders waiting at the best bid and ask prices, respectively, significantly explain future market volatility at a 5% level. A decrease in the volume of orders at the best quotes creates higher subsequent volatility. However, when the relative liquidity measures are included in the specification, $\overline{\mathrm{DEPTH1}}$ variables are no longer significant. The highest adjusted R^2 is only 26.11% even when the total depth up to the third and up to the fifth quotes are included in the analysis. That is, by including 10 depth variables in addition to our measures, we only increase the adjusted R^2 by less than 1%.

In summary, we conclude that the relative concentration of depth provision, rather than the absolute volume, reveals more information about future volatility. RLIQ has a superior in-sample predictive power compared to the standard depth measures.

5.3 Predicting further horizons

In this section, we examine the informativeness of the limit order book distribution at time τ on multiple-period-ahead volatilities. Specifically, we run the baseline regression model (2), while we calculate the dependent variable as the mid-quote volatility of the index at time $\tau + h$, with $h = 1, 2, ..., 11$, i.e., up to 165 minutes ahead. Whereas the independent variables are calculated based on the limit order book information at trading interval τ. For example, $\tau + 2$ refers to the 30-minute-ahead volatility.

Table 4 shows that the significance of the estimated coefficients as well as the predictive power of relative liquidity measures is (almost) monotonically decreasing with the prediction horizon. RLIQ^{buy} has a significant forecasting power with respect to market volatility up to 150-minutes-ahead. Moreover, the slope of the book, the relative spread, and the quote-slope significantly predict longer term volatility. However, RLIQ has a leading role in explaining longer horizon future volatility in terms of estimated standardized coefficients and t-values.

5.4 Out-of-sample tests

In this section, we evaluate the out-of-sample forecasting ability of RLIQ compared to historical realized volatility. Specifically, for a subsample of observations up to a given time interval τ, we compare the h-period-ahead squared forecast errors with the squared difference between the realized value at $\tau + h$ and the sample mean value up to time τ. To do so, we split our data into two subsample periods: T_{train} is the training period and T_{test} is the testing period with $T_{\text{train}} + T_{\text{test}} = T$, the total number of time intervals. We then re-estimate the parameters of the model in which we use the variable of interest as the predictor. Recursive estimators of

h-period-ahead forecasts are based on the sample starting from T_{train} up to $T-h$. For T_{train} equals to 400 and 350 observations, we calculate the following error terms:

$$\begin{aligned}
\varepsilon_{1,\tau+h} &= \sigma^M_{\tau+h} - \widehat{\sigma^M_{\tau+h}}, \\
\varepsilon_{2,\tau+h} &= \sigma^M_{\tau+h} - \overline{\sigma^M_{\tau}},
\end{aligned}$$

where $\sigma^M_{\tau+h}$ is the two scales realized market volatility, $\overline{\sigma^M_{\tau}}$ is the mean value of the market volatility up to time τ, and $\widehat{\sigma^M_{\tau+h}}$ is the fitted market volatilility obtained by regressing the realized volatility on the variable of interest such as RLIQ or other liquidity measures.

We evaluate the comparison by using two different metrics: the difference in mean-squared errors (ΔMSE) and the out-of-sample R^2. If the proposed measure has superior out-of-sample forecasting ability relative to the average of past data, then both of these measures will be positive. We employ the Diebold and Mariano (1995) predictive ability test (DM) to test the significance of ΔMSE. Finally, the out-of-sample R^2 is calculated as follows:

$$R^2 = 1 - \frac{\sum_{\tau=1}^{T_{\text{test}}-h} \varepsilon^2_{1,\tau+h}}{\sum_{\tau=1}^{T_{\text{test}}-h} \varepsilon^2_{2,\tau+h}}. \tag{4}$$

Table 5 reveals that forecasts based either on relative liquidity, relative spread, number of trades, or slope of the best quotes increase the predictive power relative to forecasts based only on the sample mean of past volatility. Moreover, the predictive power of the variables are decreasing almost monotonically with the prediction horizon. When $T_{\text{train}} = 400$, RLIQ^{buy} delivers an out-of-sample R^2's from 12.9% when forecasting one-period-ahead market volatility up to 5.5% when predicting 90-minutes-ahead market volatility.

On the other hand, the results are stronger for both relative spread and log quote slope when $T_{\text{train}} = 350$. Note that both $\overline{\text{relSPR}}$ and $\overline{\text{logQS}}$ are the variables that capture the liquidity at the best quotes, i.e., the tightness dimension of liquidity only. Thus as a further analysis, we examine whether including the relative liquidity, in addition to the tightness dimension of liquidity, produces better forecasts. To do so, the first forecast errors are calculated from the model where $\mathrm{RLIQ}^{\mathrm{buy}}$ and $\overline{\mathrm{relSPR}}$ ($\overline{\mathrm{logQS}}$) are the explanatory variables, whereas the second (benchmark) forecast errors are calculated from the model in which relative spread (log quote slope) is the only explanatory variable. Similarly, we repeat the analysis for two different estimation window sizes; 400 and 350 observations. The results show that including RLIQ into the analysis increases the out-of-sample R^2 by 13.9% and 11.7% in addition to using only relative spread and the log quote slope, respectively for $T_{\mathrm{train}} = 400$. Three variables together delivers an out-of-sample R^2 of over 24% when forecasting one-period-ahead market volatility relative to forecasts based only on the sample mean of realized volatility. The difference in mean-squared errors is significant at 5% up to 90 minutes ahead. Hence, we conclude that capturing both the tightness and the depth dimension of liquidity significantly increases the out-of-sample forecasting power.

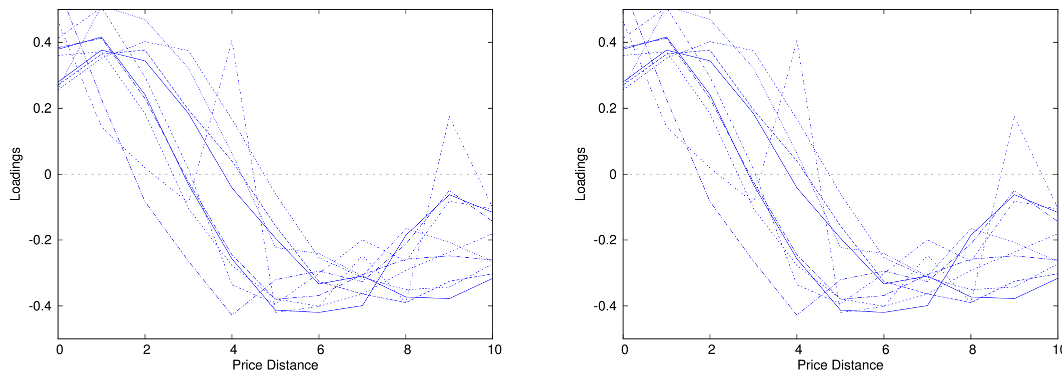
5.5 Predicting individual stock volatilities

This section examines the in-sample predictive power of a limit order book distribution over future volatility on an individual stock level. To this end, we first run the following predictive regression in a pooled data with stock fixed effects:

$$
\begin{aligned}
\sigma_{s,\tau+1} &= a_0 + a_1\sigma_{s,\tau} + a_2\mathrm{RLIQ}^{\text{ind, buy}}_{s,\tau} + a_3\mathrm{RLIQ}^{\text{ind, sell}}_{s,\tau} + \sum_{k=1}^{20} b_k T_{k,\tau} \\
&+ \sum_{s=1}^{30} c_s D_s + \text{controls} + \varepsilon_{s,\tau},
\end{aligned} \tag{5}
$$

where for stock s and interval τ, $\sigma_{s,\tau}$ is the mid-quote two scales realized volatility, $\mathrm{RLIQ}^{\text{ind, buy}}_{s,\tau}$ and $\mathrm{RLIQ}^{\text{ind, sell}}_{s,\tau}$ are the first principal components of the individual stock limit order book distributions for the buy and sell sides of the market, respectively. $T_{k,\tau}$ is the intraday dummy that equals to 1 if $k = \tau$, and D_s are stock-specific dummy variables allowing for stock fixed effects.

Figure 3 reveals that the loadings of the first principal components differ slightly from one stock to another stock and they are similar to the loadings of the first principal component of the aggregate limit order book distribution presented in Figure 2.



(a) Panel A: Buy Side

(b) Panel B: Sell Side

Figure 3: Loadings of the first principal component for the buy side of the market (Panel A) and the sell side of the market (Panel B). The loadings of a stock is presented only if the relative liquidity of the given stock is a significant predictor of 15-minute-ahead volatility.

Table 6 columns I to IV report the estimated coefficients for (5) with the corresponding t-statistics. To take into account the possible cross-sectional variations that cannot be captured by the stock fixed effects, we as well estimate the predictive regressions for each stock s separately. The summary of these results are presented in columns V to VIII.

Our main result is confirmed in these individual volatility regressions. RLIQ is negatively related to future volatility for 87% of the stocks for the buy side of the market at a 5% level. We conclude that the time-series relationship between the aggregate liquidity and market volatility is not driven by variations in a particular stock or industry, but rather it is shared by the majority of the stocks. The results reveal the asymmetry between the buy and sell sides of the market at the individual stock level as well. The sell side of the market is informative for 37% of the stocks in the individual regressions at a 5% level. Although both sides of the market are significant in the pooled regression, the estimated coefficients of the buy side are at least two times greater than the sell side.

5.6 Robustness

5.6.1 Relative liquidity: alternative measures

We define relative liquidity as the first principal component of the limit order book distribution. In this section we examine the sensitivity of the presented findings when the definition of the proposed measure, RLIQ, is changed. First, instead of using the first principal component, we consider the first three and five principal components of the aggregate limit order book distribution introduced in (1), which explain 65% and 77% of the variation, respectively. Table 7 shows that the first principal components for both buy and sell sides have the leading explanatory power for future market volatility and including the first three and five principal components, in addition to the first component, increases the adjusted R^2 by only 0.23% and 0.61%, respectively. Moreover, they do not have significant predictive power over volatility when other control variables are included.

One could easily argue in favor of the use of the empirical frequencies of orders waiting at each price distance instead of summarizing this information. Hence, as a second robustness test we consider the volume distribution for $\Delta_c = 10$ separately for each bin as predictors of volatility and exclude the last bin to avoid multicollinearity. In unreported results, we find that, by including 20 variables, instead of using only RLIQ^{buy} and $\text{RLIQ}^{\text{sell}}$, the adjusted R^2 increases only by 1.01%.

Finally, we adopt another variable selection technique, least absolute shrinkage and selection operator (LASSO, Tibshirani, 1996), to reduce the dimensionality instead of employing principal component analysis. LASSO finds the coefficients of a model by minimizing the sum of squared residuals plus an I1-norm penalty function. We reach the sparse model which corresponds to minimum mean squared errors by employing a 10-fold cross validation technique. Several conclusions arise from Figure 4. First, the estimated coefficients are negative for the bins closer to the best quotes, suggesting that higher liquidity around the best quotes are associated with lower subsequent volatility. On the other hand, the coefficients switch sign after the five best quotes, similar to the loadings of the first principal component plotted in Figure 2. The signs of the loadings of the RLIQ and those of the estimated LASSO coefficients are opposite, as expected. On the one hand, the loadings are summarizing the liquidity provision in the limit order book; the higher RLIQ, the higher the liquidity around the best quotes. On the other hand, the LASSO coefficients summarize the relationship between liquidity and volatility; the higher the liquidity around the best quotes, the lower the volatility.

Second, we see that the estimated coefficients of the sell side distribution are smaller in absolute terms compared to the buy side, suggesting that the buy side is more informative on volatility compared to the sell side of the market, in line with the results presented in Section 5. Indeed, for the sell side of the market, LASSO assigns 0 loadings to the frequency of orders waiting at the second best quotes and fourth best quotes.

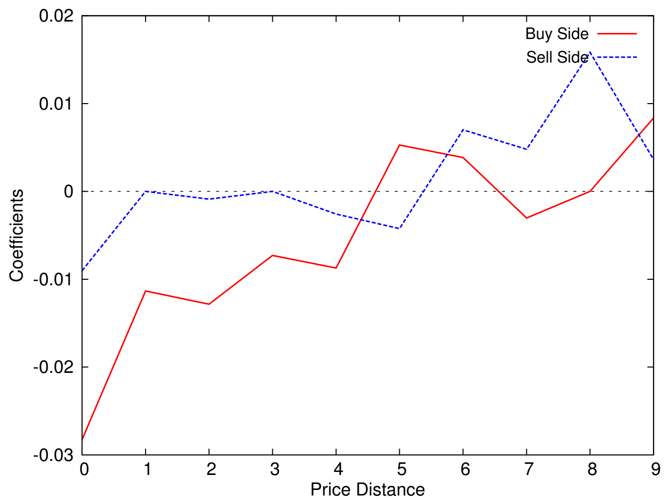


Figure 4: Estimated LASSO coefficients when $\Delta_c = 9$.

5.6.2 Further Robustness Tests

We perform five sets of additional robustness tests. First, instead of calculating our dependent variable by adopting two scales realized volatility estimator, we calculate the squared mid-quote returns of the value weighted index for each trading interval, sampled for 30 seconds. The results presented so far consider the orders to be traded up to the 10 best quotes, i.e., setting $\Delta_c = 10$ in (1). Our second set of robustness checks includes the re-calculation of RLIQ when the information up to the 20 and 30 best quotes (considering the whole book) are used. Third, instead of sampling the trading day using the 15-minute snapshots, we test the predictive power of the limit order book distribution over volatility using 30-minute intervals.

In our analysis, to proxy the aggregate level of liquidity, we first calculate RLIQ for each stock and get the cross-sectional average. Our next robustness check includes the re-calculation of the aggregate measures by using value-weighted and number-of-trades-weighted cross-sectional averages. The former weights are calculated by using the market capitalization values of the individual stocks at the end of the sample period, whereas we calculate the latter by using the daily average number of trades.

Finally, we perform a robustness test on the specification of the regression model. We re-estimate the benchmark specification in equation (2) with the log-transformed variables to allow the left-hand side of the equation to include potentially both positive and negative numbers.

The Table 8 confirms the robust relationship between the relative liquidity of the buy side of the market and future volatility. We observe that the definition of volatility, the sampling frequency, and considering the whole book instead of the first 10 best quotes do not change the results. Interestingly, we find that the sell side of the market turns to be significant when the aggregate sell side RLIQ is approximated as the value-weighted or trade-weighted average of the individual stocks. It suggests that the bigger and more actively traded stocks are the ones that are informative on future volatility.

6 Conclusion

As of today, most of the equity and derivatives exchanges around the world are either pure order driven or at least allowing limit orders in addition to the on-floor market making. The role of limit orders in trading processes expanded progressively over the last decade. This shift in trading is followed by a growing academic literature. This paper contributes to the literature, which studies the informativeness of a limit order book on future volatility. However, this is the first study that examines the predictive power of aggregate liquidity distribution over intraday market volatility. The evidence presented in this paper suggests that the state of a limit order book contains non-negligible information about the short-term aggregate price formation process. In particular, we document evidence that the distribution of quoted depth predicts both market and individual stock volatilities.

To measure the relative depth provision, we propose a new way of summarizing the distribution of liquidity in a limit order book, while taking into account the relative magnitude and the location of the quoted depth. Our summary measure relative liquidity, RLIQ, considers how liquidity is distributed in the whole book and assigns weights to the information provided by different quotes. By using high-frequency data from the Istanbul Stock Exchange, we show that RLIQ has a strong in-sample and out-of-sample predictive power with respect to market volatility, where the relationship is significant up to 75 minutes.

In a market microstructure context, information on future volatility is important because the execution probability of a limit order increases with volatility. Put differently, the probability that the current price hits the pre-determined limit price increases when volatility is higher. Hence, the relationship presented in this paper can be used to design trading strategies that may allow market participants to submit less aggressive orders and reduce execution costs.

Appendix A. Calculation of Relative Liquidity (RLIQ)

Suppose that the limit order book for stock X at 11:00am is as follows:

Order type	Volume	Limit price	Time	Best Bid	Best Ask
Sell	50,000	8.4	09:30:00	-	8.2
Buy	10,000	7.6	09:30:01	7.9	8.2
Sell	1,800	8.3	09:30:02	7.9	8.2
.					
.					
.					
Sell	3,334	8.05	10:58:17	8	8.05
Buy	25,000	8	10:58:20	8	8.05
Buy	50,000	7.9	10:58:38	8	8.05
Sell	1	8.1	10:58:50	8	8.05

The first step in the calculation of RLIQ involves the calculation of the tick-adjusted price distance Δ of each limit order in the given book relative to the best limit price:

$$\begin{aligned}\Delta_{i,\tau}^{\text{buy}} &= (p_\tau^B - p_i^{\text{buy}})/\text{tick},\\ \Delta_{i,\tau}^{\text{sell}} &= (p_i^{\text{sell}} - p_\tau^A)/\text{tick},\end{aligned}$$

where p_τ^B (p_τ^A) is the best bid (ask) price in interval τ. In this example p_τ^B=8 and p_τ^A=8.05. On the other hand, p_i^{buy} (p_i^{sell}) is the limit price of the i^{th} order.

Say the tick size is 0.05. Then we have the following price distances for each order:

Order type	Volume	Limit Price	Time	Best Bid	Best Ask	Δ
Sell	50,000	8.4	09:30:00	-	8.2	7
Buy	10,000	7.6	09:30:01	7.9	8.2	8
Sell	1,800	8.3	09:30:02	7.9	8.2	5
.					.	.
.					.	.
.					.	.
Sell	3,334	8.05	10:58:17	8	8.05	0
Buy	25,000	8	10:58:20	8	8.05	0
Buy	50,000	7.9	10:58:38	8	8.05	2
Sell	1	8.1	10:58:50	8	8.05	1

Next, we obtain of the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, .., 30$. This way, we reach the limit order book probability density function (LOB–PDF). That is:

	Buy side		**Sell side**	
Δ	Total Volume	Frequency	Total Volume	Frequency
0	78,500	0.270	68,400	0.186
1	52,575	0.181	71,602	0.194
2	58,440	0.201	54,588	0.148
3	45,579	0.156	62,068	0.168
.		.		
.		.		
.		.		
29	0	0.000	0	0.000
30	0	0.000	0	0.000

By repeating the procedure for each of the time interval τ, we end up a time series of frequencies for each price distance Δ. Hence, the LOB–PDF of stock X for a given day and for the sell side of the market looks as follows,

				Frequencies					
Time/ Δ	**0**	**1**	**2**	**3**	.	.	.	**29**	**30**
10:00	0.212	0.259	0.182	0.133	.	.	.	0.000	0.000
10:15	0.214	0.249	0.183	0.120	.	.	.	0.000	0.000
10:30	0.180	0.243	0.184	0.122	.	.	.	0.000	0.000
10:45	0.194	0.230	0.160	0.124	.	.	.	0.000	0.000
11:00	**0.186**	**0.194**	**0.148**	**0.168**	.	.	.	**0.000**	**0.000**
.					.	.	.		
.					.	.	.		
.					.	.	.		
16:30	0.213	0.223	0.146	0.112	.	.	.	0.000	0.000
16:45	0.213	0.224	0.156	0.122	.	.	.	0.000	0.000
17:00	0.188	0.240	0.171	0.118	.	.	.	0.000	0.000

The avgPDF is obtained as the equally-weighted cross-sectional average of the individual LOB–PDFs. In other words, we obtain the previous table for all of the stocks in our sample and calculate the cross-sectional average of frequencies to have a distribution of the market for a given delta. RLIQ_τ is the summary measure of this aggregate limit order book distribution obtained as the first principal component of the frequency of orders waiting at different price levels at interval τ.

References

Ahn, H., K. Bae, and K. Chan, 2001, Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong, *Journal of Finance* 56, 767–788.

Ait-Sahalia, Y., P. Mykland, and L. Zhang, 2005, How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies* 18, 351–416.

Ait-Sahalia, Y., P. Mykland, and L. Zhang, 2011, Ultra high frequency volatility estimation with dependent microstructure noise, *Journal of Econometrics* 160, 160–175.

Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Bandi, F., M., and R. Russell, J., 2009, Microstructure Noise, Realized Variance, and Optimal Sampling, *The Review of Economic Studies* 75, 339–369.

Bollerslev, T., and I. Domowitz, 1993, Trading Patterns and Prices in the Interbank Foreign Exchange Market., *Journal of Finance* 48, 1421–1443.

Burdett, K., and M. O'Hara, 1987, Building blocks: An introduction to block trading, *Journal of Banking and Finance* 11, 193–212.

Cao, C., O. Hansch, and X. Wang, 2008, Order placement strategies in a pure limit order book market., *The Journal of Financial Research* 31, 113–140.

Cao, C., O. Hansch, and X. Wang, 2009, The Information content of an open Limit-Order book, *The Journal of Future Markets* 29, 16–41.

Coppejans, M., I. Domowitz, and A. Madhavan, 2001, Liquidity in an Automated Auction, Working paper, BlackRock; Barclays Global Investors.

Delattre, S., Y. Robert, C., and M. Rosenbaum, 2013, Estimating the efficient price from the order flow: a Brownian Cox process approach, Working paper.

Diebold, F. X., and R. S. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Domowitz, I., O. Hansch, and X. Wang, 2005, Liquidity commonality and return co-movement, *Journal of Financial Markets* 8, 351–376.

Duong, H., and P. Kalev, 2008, Order Book Slope and Price Volatility, Working paper, Monash University.

Ekinci, C., 2008, *Intraday Liquidity in the Istanbul Stock Exchange,* John Wiley & Sons, chap. 6, pp. 77–94.

Ellul, A., C. Holden, P. Jain, and R. Jennings, 2007, Order dynamics: Recent evidence from the NYSE, *Journal of Empirical Finance* 14, 636–661.

Foucault, T., 1999, Order flow Composition and Trading Costs in a Dynamic Limit Order Market, *Journal of Financial Markets* 2, 99–134.

Foucault, T., O. Kadan, and E. Kandel, 2005, Limit Order Book as a Market for Liquidity, *Review of Financial Studies* 18, 1171–1217.

Foucault, T., S. Moinas, and E. Theissen, 2007, Does Anonymity Matter in Electronic Limit Order Markets?, *Review of Financial Studies* 20, 1707–1747.

Ghysels, E., and A. Sinko, 2011, Volatility forecasting and microstructure noise, *Journal of Econometrics* 160, 257–271.

Goettler, R., C. Parlour, and U. Rajan, 2005, Equilibrium in a Dynamic Limit Order Market, *Journal of Finance* 60, 2149–2192.

Goettler, R., C. Parlour, and U. Rajan, 2009, Informed traders and limit order markets, *Journal of Financial Economics* 93, 67–87.

Griffiths, M., B. Smith, A. Turnbull, and R. White, 2000, The costs and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65–88.

Handa, P., and R. Schwartz, 1996, Limit Order Trading, *Journal of Finance* 51, 1835–1861.

Hansen, P., R., and A. Lunde, 2006, Realized Variance and Market Microstructure Noise, *Journal of Business & Economic Statistics* 24, 127–161.

Hasbrouck, J., and J. Seppi, D., 2001, Common Factors in Prices, Order Flows, and Liquidity, *Journal of Financial Economics* 59, 383–411.

Jones, CM., G. Kaul, and ML. Lipson, 1994, Transactions, volume, and volatility, *Review of Financial Studies* 7, 631–651.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun, 2011, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, Working Paper, University of Maryland.

Koksal, B., 2012, An analysis of intraday patterns and liquidity on the Istanbul stock exchange, Working paper Central Bank of Turkey.

Marshall, B., R., 2006, Liquidity and stock returns: Evidence from a pure order-driven market using a new liquidity proxy, *International Review of Financial Analysis* 15, 21–38.

Naes, R., and J. Skjeltorp, 2006, Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market, *Journal of Financial Markets* 9, 408–432.

Parlour, C., 1998, Price Dynamics in Limit Order Markets, *Review of Financial Studies* 11, 789–816.

Pascual, R., and D. Veredas, 2009, What Pieces of Limit Order Book Information Matter in Explaining Order Choice by Patient and Impatient Traders, *Quantitative Finance* 9, 527–545.

Pascual, R., and D. Veredas, 2010, Does the Open Limit Order Book Matter in Explaining Informational Volatility?, *Journal of Financial Econometrics* 8, 57–87.

Poon, S., and C. W. J. Granger, 2003, Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature* 16, 478–539.

Porter, R., B., 2008, *The Multiple Dimensions of Marketwide Liquidity: Implications for Asset Pricing,* John Wiley & Sons, chap. 21, pp. 387–406.

Ranaldo, A., 2004, Order aggressiveness in limit order book markets, *Journal of Financial Markets* 7, 53–74.

Schwert, G. W., 1989, Why does stock market volatility change over time, *Journal of Finance* 44, 1115–1153.

Tibshirani, R., 1996, Regression Shrinkage and Selection via the LASSO, *Journal of the Royal Statistical Society* 58, 267–288.

Valenzuela, M., and I. Zer, 2013, Competition, signaling and non–walking through the book: Effects on order choice, *Journal of Banking and Finance* 37, 5421–5435.

Zhang, L., P.A. Mykland, and Y. Ait-Sahalia, 2005, A tale of two time scales: determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394–1411.

Zhou, B., 1996, High-Frequency Data and Volatility in Foreign-Exchange Rates, *Journal of Business & Economic Statistics* 14, 45–52.

Table 1: Summary Statistics: The Limit Order Book Distribution

Descriptive statistics for the empirical limit order book distributions for both sides the market. The mean, variance, skewness, and the fractions of number of shares accumulated up to a given price distance Δ are reported. The first column shows the summary statistics of the limit order book distribution which is obtained by averaging across intervals, days, and stocks. The last four columns report the statistics for four limit order book distributions (averaged across stocks) at 10:00 (beginning of the day), 12:00 (end of the morning session), 14:15 (beginning of the afternoon session) and 17:00 (end of the trading day).

		uncond.	**10:00**	**12:00**	**14:15**	**17:00**
Buy side	mean	3.43	3.64	3.32	3.41	3.42
	variance	18.42	20.06	17.67	17.83	17.52
	skewness	2.41	2.34	2.60	2.33	2.35
	up to 1 Δ	0.40	0.38	0.40	0.41	0.41
	up to 3 Δ	0.68	0.66	0.69	0.68	0.68
	up to 5 Δ	0.82	0.81	0.84	0.82	0.82
	up to 10 Δ	0.93	0.92	0.94	0.93	0.93
	up to 20 Δ	0.99	0.99	0.99	0.99	0.99
	up to 30 Δ	1.00	1.00	1.00	1.00	1.00
Sell side	mean	4.63	4.68	4.64	4.56	4.73
	variance	25.16	27.51	25.77	23.73	24.20
	skewness	1.84	1.83	1.89	1.77	1.74
	up to 1 Δ	0.30	0.31	0.29	0.30	0.28
	up to 3 Δ	0.55	0.56	0.55	0.55	0.53
	up to 5 Δ	0.72	0.73	0.72	0.72	0.70
	up to 10 Δ	0.88	0.87	0.88	0.89	0.88
	up to 20 Δ	0.98	0.98	0.98	0.98	0.98
	up to 30 Δ	1.00	1.00	1.00	1.00	1.00

Table 2: Predictability Regressions

Estimated coefficients of the regression model defined in (2). The dependent variable is the 15-minutes-ahead market volatility, $\sigma^M_{\tau+15\text{min}}$, calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100). RLIQ^{buy} ($\text{RLIQ}^{\text{sell}}$) is the first principal components of the respective aggregate limit order book distributions for the buy (sell) side as outlined in Section 3.2. All of the control variables are constructed as the cross-sectional average of the corresponding individual stock measures. SLOPE is the slope of the limit order book, relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, DHW is the Domowitz, Hansch, and Wang (2005) illiquidity measure. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. Var.: $\sigma^M_{\tau+15\text{min}}$	**I**	**II**	**III**	**IV**	**V**	**VI**
RLIQ^{buy}	-0.044		-0.035	-0.031	-0.029	-0.027
	(-7.38)		(-7.47)	(-7.16)	(-6.29)	(-5.70)
$\text{RLIQ}^{\text{sell}}$	-0.028		-0.022	-0.017	-0.010	-0.010
	(-4.25)		(-4.14)	(-3.45)	(-1.75)	(-1.73)
$\overline{\text{SLOPE}}_\tau$				0.009	0.016	0.014
				(1.45)	(2.90)	(2.38)
$\overline{\text{relSPR}}_\tau$				0.025	0.013	0.012
				(4.89)	(1.77)	(1.54)
$\overline{\text{NT}}_\tau$				0.007		0.007
				(1.27)		(1.28)
$\overline{\text{AQ}}_\tau$				0.000		0.002
				(0.10)		(0.36)
$\overline{\text{AMR}}_\tau$					0.001	0.001
					(0.33)	(0.45)
$\overline{\text{logQS}}_\tau$					0.024	0.025
					(2.13)	(2.16)
$\overline{\text{DHW}}_\tau$					0.001	0.001
					(0.17)	(0.20)
σ^M_τ		0.037	0.023	0.015	0.015	0.012
		(6.28)	(5.04)	(3.32)	(3.39)	(2.51)
constant	0.240	0.238	0.233	0.228	0.243	0.240
	(13.08)	(13.31)	(14.15)	(13.48)	(14.14)	(14.09)
adj. R^2(%)	21.95	17.25	25.15	28.48	29.45	29.53

Table 3: RLIQ vs standard depth measures

Comparison of in-sample predictive power of RLIQ with respect to the depth measures, i.e., the quoted volume of orders waiting at a given threshold. The dependent variable is the market volatility, $\sigma^M_{\tau+15min}$ calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100). RLIQ$^{\text{buy}}$ (RLIQ$^{\text{sell}}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side as outlined in Section 3.2. $\overline{\text{DEPTH}i}^{\text{buy}}_{\tau}$ is the quoted depth at price distance i. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. Var.: $\sigma^M_{\tau+15\text{min}}$	**I**	**II**	**III**	**IV**	**V**
RLIQ$^{\text{buy}}$	-0.035		-0.032	-0.033	-0.036
	(-7.47)		(-7.02)	(-6.84)	(-7.04)
RLIQ$^{\text{sell}}$	-0.022		-0.018	-0.016	-0.008
	(-4.14)		(-3.03)	(-2.54)	(-1.10)
$\overline{\text{DEPTH1}}^{\text{buy}}_{\tau}$		-0.012	-0.003	-0.001	0.000
		(-2.26)	(-0.66)	(-0.21)	(0.00)
$\overline{\text{DEPTH1}}^{\text{sell}}_{\tau}$		-0.011	-0.006	-0.003	-0.003
		(-2.50)	(-1.55)	(-0.51)	(-0.46)
$\overline{\text{DEPTH2}}^{\text{buy}}_{\tau}$				-0.007	-0.008
				(-1.11)	(-1.34)
$\overline{\text{DEPTH2}}^{\text{sell}}_{\tau}$				0.001	0.000
				(0.14)	(0.06)
$\overline{\text{DEPTH3}}^{\text{buy}}_{\tau}$				-0.001	0.010
				(-0.15)	(1.29)
$\overline{\text{DEPTH3}}^{\text{sell}}_{\tau}$				-0.001	-0.011
				(-0.10)	(-1.42)
$\overline{\text{DEPTH4}}^{\text{buy}}_{\tau}$					-0.017
					(-2.67)
$\overline{\text{DEPTH4}}^{\text{sell}}_{\tau}$					0.012
					(1.38)
$\overline{\text{DEPTH5}}^{\text{buy}}_{\tau}$					-0.005
					(-0.84)
$\overline{\text{DEPTH5}}^{\text{sell}}_{\tau}$					0.002
					(0.23)
σ^M_{τ}	0.023	0.032	0.022	0.022	0.021
	(5.04)	(6.31)	(4.96)	(5.00)	(4.73)
constant	0.233	0.252	0.240	0.241	0.247
	(14.15)	(14.01)	(13.69)	(13.61)	(13.79)
adj. R^2(%)	25.15	19.94	25.32	25.08	26.11

Table 4: Predictability Regressions–Further Horizons

Estimated coefficients of the regression model defined in (2). The dependent variable is the market volatility, $\sigma^M_{\tau+h}$ calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100) in period $\tau + h$ for $h = 1, 2, ..., 11$. RLIQ$^{\text{buy}}$ (RLIQ$^{\text{sell}}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side as outlined in Section 3.2. All of the control variables are constructed as the cross-sectional average of the corresponding individual stock measures. SLOPE is the slope of the limit order book, relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, DHW is the Domowitz, Hansch, and Wang (2005) illiquidity measure. In Panel A for every time horizon, we report the "simple" regressions, where the relative liquidity measures along with the lagged volatility and interval dummies are used as regressors. On the other hand Panel B reports the results when all of the control variables are included in the regression equation. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. var.: $\sigma^M_{\tau+h}$	**Panel A: "simple" regressions**							**Panel B: multiple regressions**						
	0–15	**15–30**	**30–45**	.	**120–135**	**135–150**	**150–165**	**0–15**	**15–30**	**30–45**	.	**120–135**	**135–150**	**150–165**
RLIQ$^{\text{buy}}$	-0.035	-0.032	-0.031	.	-0.030	-0.028	-0.025	-0.027	-0.023	-0.023	.	-0.03	-0.029	-0.017
	(-7.47)	(-5.97)	(-5.28)	.	(-3.95)	(-3.21)	(-2.77)	(-5.70)	(-5.33)	(-4.48)	.	(-3.87)	(-3.92)	(-1.98)
RLIQ$^{\text{sell}}$	-0.022	-0.025	-0.023	.	-0.016	-0.014	-0.015	-0.010	-0.011	-0.008	.	0.00	-0.003	0.003
	(-4.14)	(-3.78)	(-3.12)	.	(-1.92)	(-1.66)	(-1.85)	(-1.73)	(-2.26)	(-1.22)	.	(0.16)	(-0.38)	(0.33)
$\overline{\text{SLOPE}}_\tau$				.				0.014	0.021	0.019	.	0.01	0.012	0.010
				.				(2.38)	(3.04)	(2.67)	.	(1.44)	(1.39)	(1.34)
$\overline{\text{relSPR}}_\tau$				.				0.012	0.015	0.015	.	0.03	0.045	0.010
				.				(1.54)	(2.44)	(2.30)	.	(3.06)	(3.79)	(1.07)
$\overline{\text{NT}}_\tau$				.				0.007	0.010	0.007	.	0.00	0.001	0.011
				.				(1.28)	(1.59)	(1.13)	.	(0.25)	(0.08)	(1.42)
$\overline{\text{AQ}}_\tau$				.				0.002	0.004	0.002	.	0.00	-0.009	0.011
				.				(0.36)	(0.89)	(0.54)	.	(0.04)	(-1.46)	(1.54)
$\overline{\text{AMR}}_\tau$				.				0.001	0.007	-0.001	.	0.00	-0.001	0.001
				.				(0.45)	(5.98)	(-0.85)	.	(0.20)	(-0.22)	(0.69)
$\overline{\text{logQS}}_\tau$				.				0.025	0.030	0.031	.	0.00	-0.017	0.028
				.				(2.16)	(4.13)	(3.85)	.	(0.29)	(-1.23)	(2.30)
$\overline{\text{DHW}}_\tau$				.				0.001	0.001	0.003	.	0.02	0.015	0.013
				.				(0.20)	(0.21)	(0.64)	.	(2.08)	(1.94)	(2.14)
σ^M_τ	0.023	0.016	0.006	.	0.005	0.011	0.011	0.012	0.001	-0.007	.	0.00	0.007	-0.006
	(5.04)	(4.17)	(1.24)	.	(1.00)	(1.66)	(1.62)	(2.51)	(0.32)	(-1.26)	.	(-0.76)	(1.02)	(-0.88)
constant	0.233	0.244	0.242	.	0.244	0.244	0.243	0.240	0.255	0.255	.	0.25	0.245	0.237
	(14.15)	(13.44)	(12.97)	.	(12.48)	(12.99)	(13.00)	(14.09)	(14.51)	(13.87)	.	(13.66)	(13.68)	(12.65)
adj. R^2(%)	25.15	20.93	16.73	.	12.61	13.36	14.01	29.53	28.03	23.57	.	19.20	20.00	21.81

Table 5: Out-of-Sample Forecasting Evaluation

Out-of-sample forecasting experiment results. The h-period-ahead forecast error is obtained as the difference between the realized volatility at $\tau + h$ and the fitted value of the predictive regression estimated up to time τ with the variable of interest listed in the first column is used as a regressor. On the other hand, the competing error is calculated from the sample mean volatility up to time interval τ. The dependent variable is the market volatility, $\sigma^M_{\tau+h}$ calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100) in period $\tau + h$ for $h = 1, 2, ..., 6$. RLIQ^{buy} ($\text{RLIQ}^{\text{sell}}$) is the first principal component of the aggregate limit order book distribution for the buy (sell) side as outlined in Section 3.2. $\overline{\text{relSPR}}_\tau$, $\overline{\text{NT}}_\tau$, and $\overline{\text{logQS}}_\tau$ are the cross-sectional averages of the relative spread, number of trades, and log quote slope respectively. Although we examine the out-of-sample forecasting power of all of the control variables introduced in Section 4.3, for the sake of brevity we report only the ones with significant forecasting power. The out-of-sample $R^2(\%)$ and the difference in mean-squared errors (ΔMSEx1000) are reported. Finally, DM denotes the Diebold and Mariano (1995) predictive ability test. Panels A and B report the results when the training period is set to 400 and 350 observations, respectively.

Forecasting variable		**0–15min**	**15–30min**	**30–45min**	**45–60min**	**60–75min**	**75–90min**
Panel A: Training Period: 400 obs.							
RLIQ^{buy}	Out-of-sample $R^2(\%)$	12.86	10.13	9.20	8.25	7.08	5.50
	ΔMSE	1.91	1.50	1.36	1.23	1.06	0.83
	DM t-stat	2.68	2.45	2.74	2.46	2.35	1.86
$\overline{\text{relSPR}}_\tau$	Out-of-sample $R^2(\%)$	10.85	9.95	8.94	9.13	9.48	10.57
	ΔMSE	1.61	1.47	1.32	1.36	1.42	1.59
	DM t-stat	2.24	2.18	1.94	1.94	1.99	2.30
$\overline{\text{NT}}_\tau$	Out-of-sample $R^2(\%)$	12.89	8.32	0.69	-0.54	0.38	2.16
	ΔMSE	1.91	1.23	0.10	-0.08	0.06	0.32
	DM t-stat	2.46	2.00	0.31	-0.49	0.25	0.87
$\overline{\text{logQS}}_\tau$	Out-of-sample $R^2(\%)$	14.83	12.59	10.57	10.54	10.04	11.83
	ΔMSE	2.20	1.86	1.56	1.57	1.50	1.78
	DM t-stat	2.92	2.66	2.50	2.43	2.24	2.41
Panel B: Training Period: 350 obs.							
RLIQ^{buy}	Out-of-sample $R^2(\%)$	11.23	8.70	7.95	7.19	6.32	4.93
	ΔMSE	1.80	1.40	1.28	1.16	1.03	0.80
	DM t-stat	2.76	2.44	2.69	2.46	2.37	1.90
$\overline{\text{relSPR}}_\tau$	Out-of-sample $R^2(\%)$	13.35	12.38	11.66	11.68	11.89	12.59
	ΔMSE	2.14	1.99	1.88	1.89	1.93	2.05
	DM t-stat	2.98	2.93	2.75	2.73	2.74	3.03
$\overline{\text{NT}}_\tau$	Out-of-sample $R^2(\%)$	11.11	6.91	0.73	-0.41	0.03	1.24
	ΔMSE	1.78	1.11	0.12	-0.07	0.00	0.20
	DM t-stat	2.43	1.97	0.35	-0.44	0.03	0.60
$\overline{\text{logQS}}_\tau$	Out-of-sample $R^2(\%)$	15.67	12.89	11.10	10.66	11.34	12.56
	ΔMSE	2.51	2.07	1.79	1.73	1.84	2.05
	DM t-stat	3.60	3.22	3.12	2.92	2.95	3.00

Table 6: Individual Stock Predictability Regressions

Estimated coefficients of the regression model defined in (5). $\text{RLIQ}^{\text{ind, buy}}$ ($\text{RLIQ}^{\text{ind, sell}}$) is the first principal component of the individual stock limit order book distribution, for the buy (sell) side. In a given trading interval τ, SLOPE is the slope of the limit order book, relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001). Finally, DHW is the Domowitz, Hansch, and Wang (2005) illiquidity measure. All of the explanatory variables are standardized. The dependent variable is $\sigma_{\tau+1}$, which is the TSRV volatility calculated using the mid-quotes of the orders originated in interval $\tau+1$ (multiplied by 100). Panel A shows the results for the pooled regression. t-statistics based on cluster robust standard errors on stock level are reported. Panel B summarizes the results when the model is estimated for each stock separately. The cross-sectional median of the estimated significant coefficients at a 5% level is reported. In brackets, first, the percentage of the stocks with a significant coefficient at a 5% level and second, the percentage of the positive estimates, are reported. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals. For the sake of brevity, the estimated coefficients of the intraday dummies and stock fixed effects are omitted.

dep. var.: $\sigma_{s,\tau+1}$	**Panel A: Pooled regression**				**Panel B: Summary of individual regressions**			
	I	II	III	IV	V	VI	VII	VIII
$\text{RLIQ}^{\text{ind,buy}}$	-0.057	-0.058	-0.058	-0.055	-0.065	-0.057	-0.055	-0.052
	(-13.07)	(-12.95)	(-12.60)	(-11.94)	[87/0]	[87/0]	[87/0]	[80/0]
$\text{RLIQ}^{\text{ind,sell}}$	-0.024	-0.026	-0.019	-0.019	-0.039	-0.046	-0.048	-0.042
	(-4.99)	(-5.36)	(-3.56)	(-3.66)	[37/9]	[33/0]	[27/0]	[27/0]
SLOPE_τ		-0.019	0.031	0.022		-0.057	0.054	0.039
		(-2.47)	(4.47)	(3.58)		[43/8]	[30/78]	[20/67]
relSPR_τ		0.040	-0.001	-0.007		0.088	-0.104	-0.164
		(4.09)	(-0.14)	(-0.67)		[37/91]	[30/33]	[23/14]
NT_τ		0.037		0.040		0.048		0.057
		(9.70)		(11.18)		[57/100]		[57/100]
AQ_τ		-0.005		0.001		0.023		0.036
		(-1.02)		(0.20)		[17/60]		[17/100]
AMR_τ			0.000	0.003			0.007	0.011
			(-0.02)	(0.93)			[20/50]	[20/50]
logQS_τ			0.091	0.094			0.112	0.126
			(10.13)	(10.22)			[53/100]	[50/100]
DHW_τ			0.013	0.014			0.055	0.055
			(2.61)	(2.79)			[33/90]	[33/90]
σ_τ	0.082	0.055	0.064	0.042	0.083	0.068	0.065	0.060
	(16.19)	(9.59)	(14.29)	(7.62)	[97/100]	[50/100]	[77/100]	[37/100]
constant	0.565	0.553	0.558	0.541	0.577	0.571	0.580	0.571
	(26.41)	(25.93)	(25.06)	(23.57)	[100/100]	[100/100]	[100/100]	[100/100]
adj. R^2(%)	13.24	14.75	15.50	15.95	10.27	13.13	14.08	14.95

Table 7: One Period Ahead Predictive Power–Principal Components

In-sample predictive power of other principal components over 15-minute-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the mid-quote volatility of the value-weighted index (multiplied by 100). PCi is the i^{th} principal component of the aggregate limit order book distribution for the buy (sell) side as outlined in Section 3.2. Columns I–III report the estimated coefficients of the principal components when the lagged volatility and intraday dummies are included in the specification as control variables. In columns IV to VI the other control variables defined in Table 2 are included in addition. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. Var.: $\sigma^M_{\tau+15\text{min}}$	**I**	**II**	**III**	**IV**	**V**	**VI**
PC1^{buy} (RLIQ^{buy})	-0.035	-0.034	-0.035	-0.027	-0.028	-0.029
	(-7.47)	(-7.27)	(-7.42)	(-5.70)	(-5.63)	(-5.85)
PC1^{sell} ($\text{RLIQ}^{\text{sell}}$)	-0.022	-0.003	-0.002	-0.010	-0.001	-0.001
	(-4.14)	(-0.86)	(-0.73)	(-1.73)	(-0.13)	(-0.42)
PC2^{buy}		-0.002	-0.002		-0.001	-0.001
		(-0.54)	(-0.57)		(-0.40)	(-0.25)
PC2^{sell}		-0.022	-0.005		-0.010	-0.004
		(-3.73)	(-1.51)		(-1.71)	(-1.13)
PC3^{buy}		0.003	0.008		0.001	0.007
		(0.78)	(2.09)		(0.21)	(1.83)
PC3^{sell}		0.007	0.022		(0.003)	(0.012)
		(2.14)	(3.77)		(0.89)	(2.03)
PC4^{buy}			-0.003			0.001
			(-0.81)			(0.26)
PC4^{sell}			-0.008			-0.004
			(-2.50)			(-1.17)
PC5^{buy}			0.000			0.002
			(0.07)			(0.50)
PC5^{sell}			0.002			0.003
			(0.59)			(0.82)
$\overline{\text{SLOPE}}_\tau$				0.014	0.014	0.016
				(2.38)	(2.37)	(2.69)
$\overline{\text{relSPR}}_\tau$				0.012	0.013	0.013
				(1.54)	(1.63)	(1.68)
$\overline{\text{NT}}_\tau$				0.007	0.007	0.007
				(1.28)	(1.26)	(1.35)
$\overline{\text{AQ}}_\tau$				0.002	0.001	0.001
				(0.36)	(0.34)	(0.15)
$\overline{\text{AMR}}_\tau$				0.001	0.001	0.001
				(0.45)	(0.42)	(0.34)
$\overline{\text{logQS}}_\tau$				0.025	0.024	0.022
				(2.16)	(1.94)	(1.85)
$\overline{\text{DHW}}_\tau$				0.001	0.00	0.001
				(0.20)	(0.07)	(0.25)
σ^M_τ	0.023	0.020	0.021	0.012	0.010	0.011
	(5.04)	(4.82)	(4.54)	(2.51)	(2.48)	(2.34)
adj. R^2(%)	25.15	25.38	25.76	29.53	29.27	29.42

Table 8: Robustness

Columns I and II present the results when the dependent variable is the realized volatility. Columns III to VI report the results when Δ_c is equal to 20 and 30. The following two columns show the results when the sampling period is 30 minutes instead of 15 minutes. In columns XI to XII we report the results when the explanatory variables are aggregated via value-weighted and trade-weighted cross-sectional averages. Finally, the last two columns present the estimated coefficients for the log-transformed variables. All of the explanatory variables are standardized. In all of the specifications t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and for the sake of brevity, the estimated coefficients of the intraday dummies are omitted. All of the variables are defined in Table 2.

	$\sigma^M_{\tau+15\text{min,std}}$		$\Delta_c = 20$		$\Delta_c = 30$		30–min sampling		value–weighted		trade–weighted		log transform.	
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
RLIQ^{buy}	-0.041	-0.033	-0.033	-0.026	-0.032	-0.025	-0.053	-0.043	-0.034	-0.025	-0.038	-0.029	-0.031	-0.024
	(-6.84)	(-4.95)	(-7.71)	(-5.65)	(-7.43)	(-5.50)	(-6.26)	(-6.22)	(-7.23)	(-5.09)	(-7.33)	(-5.19)	(-7.33)	(-5.98)
$\text{RLIQ}^{\text{sell}}$	-0.024	-0.009	-0.017	-0.007	-0.016	-0.006	-0.036	-0.012	-0.023	-0.012	-0.027	-0.017	-0.022	-0.008
	(-3.73)	(-1.38)	(-3.04)	(-1.12)	(-2.83)	(-0.91)	(-3.35)	(-1.67)	(-4.53)	(-2.32)	(-4.81)	(-3.03)	(-3.87)	(-1.55)
$\overline{\text{SLOPE}}_\tau$		0.017		0.015		0.014		0.034		0.019		0.019		0.015
		(2.16)		(2.51)		(2.39)		(3.20)		(2.34)		(2.50)		(2.70)
$\overline{\text{relSPR}}_\tau$		0.014		0.012		0.013		0.025		-0.017		-0.015		0.011
		(1.57)		(1.58)		(1.67)		(2.49)		(-1.69)		(-1.52)		(1.43)
$\overline{\text{NT}}_\tau$		0.006		0.006		0.006		0.006		0.010		0.010		0.006
		(0.77)		(1.06)		(1.15)		(0.62)		(2.30)		(2.26)		(0.99)
$\overline{\text{AQ}}_\tau$		0.004		0.003		0.003		0.011		-0.003		-0.003		-0.001
		(0.69)		(0.74)		(0.77)		(1.52)		(-0.84)		(-0.92)		(-0.26)
$\overline{\text{AMR}}_\tau$		0.003		0.002		0.002		0.009		0.001		0.000		0.001
		(0.61)		(0.65)		(0.58)		(1.71)		(0.26)		(0.06)		(0.44)
$\overline{\text{logQS}}_\tau$		0.030		0.022		0.022		0.056		0.046		0.046		0.029
		(2.24)		(1.89)		(1.86)		(4.03)		(3.55)		(3.61)		(2.49)
$\overline{\text{DHW}}_\tau$		0.001		0.001		0.001		0.005		-0.001		-0.002		-0.002
		(0.12)		(0.21)		(0.16)		(0.72)		(-0.16)		(-0.39)		(-0.33)
σ^M_τ	0.031	0.018	0.022	0.012	0.023	0.012	0.033	0.004	0.024	0.013	0.023	0.012	0.025	0.015
	(5.35)	(2.90)	(4.83)	(2.56)	(4.90)	(2.58)	(4.25)	(0.51)	(5.09)	(2.68)	(4.94)	(2.45)	(5.45)	(3.15)
constant	0.300	0.306	0.234	0.237	0.236	0.238	0.382	0.400	0.230	0.238	0.230	0.238	0.232	0.241
	(14.50)	(14.93)	(14.30)	(13.78)	(14.33)	(13.88)	(15.12)	(16.47)	(14.24)	(14.38)	(14.05)	(14.30)	(14.40)	(14.28)
adj. $R^2(\%)$	25.96	29.71	25.82	29.49	25.55	29.33	31.21	41.71	25.32	28.56	25.61	29.11	24.55	28.81